May 1, 1981
T. Rindfleisch
P. Friedland
J. Clayton

THE GENET GUEST SERVICE ON SUMEX

The MOLGEN project at Stanford has focused on applications of artificial intelligence and symbolic computation to the field of molecular biology. The research began in 1975 and is currently in the first year of a three year grant renewal. In early 1980 it was realized that some of the systems developed by MOLGEN were of direct utility to many scientists in the domain. Accordingly, with the cooperation of the SUMEX-AIM staff and close coordination with the AIM Executive Committee, it was decided in February 1980 to provide a carefully limited guest service for the community use of such systems.

There were two major reasons for the establishment of this guest service, which took the form of the GENET account on SUMEX. The first was to broaden MOLGEN's base of scientist collaborators, to find molecular biologists at institutions other than Stanford who could contribute actively to our knowledge-based approach to problem solving. The second was to introduce a generally computer-naive community to the benefits of resource sharing provided by a system like SUMEX, with the hope of serving as a model for an eventual resource for molecular biology.

We believe that we have succeeded in these two goals. Many of our GENET guests have become active collaborators in core MOLGEN research. These collaborators include Professor Allan Maxam at Harvard Medical School, Dr. Walter Goad at Los Alamos, Dr. Richard Roberts at Cold Spring Harbor, Dr. William Pearson at Johns Hopkins, Drs. Walter Bodmer, Julia Bodmer, and Robert Kamen at the Imperial Cancer Research Fund, Professor Fred Blattner at Wisconsin, Dr. Andrew Taylor at University of Oregon, and Dr. Dan Davison of SUNY-Stonybrook. We are also pleased by the numerous comments SUMEX has received from GENET users praising the user-sensitive nature of the resource, especially in comparison to typical university computer centers.

GENET has been important both for MOLGEN and for the national community of molecular biology. It has ensured a steady flow of ideas for the artificial intelligence research that is core to both the MOLGEN grant and the SUMEX-AIM mission. It has also provided a useful service to an international community that is not readily available elsewhere.

GENET GUEST COMMUNITY MANAGEMENT

Our decision to support the GENET guest experiment and our approach to doing so within the SUMEX-AIM resource has been reviewed and approved both by the AIM Executive Committee and by the Initial Review Group/National Advisory Research Resources Council in the course of the peer review of our pending SUMEX renewal application. We have tried to manage the GENET guest experiment in such a way that we maintain the "friendly" interface of the SUMEX-AIM resource for molecular biologists unfamiliar with computers while taking appropriate steps so that GENET usage does not detract from on-going AI research and so that we assure prudent administration SUMEX as an NIH-BRP resource. The key elements in our management approach include:

1) Controlled announcement of the GENET opportunity -- Beginning in February 1980, the availability of GENET services was announced, primarily by talks at professional conferences with accompanying program demonstrations. We decided against publishing "blanket" announcements in professional journals in order to maintain a very high standard of collaborator interest and scientific expertise within the limited group we could serve with available SUMEX resources.

2) Close coordination with the AIM Executive Committee -- We kept the AIM Executive Committee apprised of plans for the GENET experiment and of progress and growth of the community. At the August 1980 AIM Workshop meeting of the Executive Committee, Professor L. Kedes of the MOLGEN project made a presentation on the status of GENET. The Executive Committee approved continuation of the GENET service but because of the significant growth in the number of GENET users and their consumption of CPU resources, a limit of two simultaneous GENET jobs was placed on the community. The Executive Committee also approved the concept of a proposed Molecular Biology Computing Resource related to but separate from the existing SUMEX resource.

3) Careful control of GENET usage -- We have closely monitored the very rapid growth in GENET usage of SUMEX (see data below). With Executive Committee advice and in cooperation with the MOLGEN project personnel managing the GENET community, we have instituted several successively stringent controls on GENET users:

   a) All GENET users run out of the same directory so scheduler control limits are enforced to hold GENET usage as a whole down relative to that of AI research projects during heavy loads.

   b) The GENET directory has been intentionally limited in disk space allocation so that large numbers of files cannot be retained.

   c) Starting in October 1980, a limit of two simultaneous logged-in GENET jobs was placed on the community.

d) Starting in December 1980, a policy statement was issued restricting GENET use to academic collaborators. MOLGEN project management informed industrial collaborators that they could no longer use the GENET facility and actively monitored adherence to this policy. Previously, valuable feedback had been obtained from a small group of industrial collaborators for MOLGEN AI program development. However, with the rapid growth of the highly competitive molecular genetics industry, there was no way we could adequately control industrial users consistent with SUMEX's status as a federally funded national resource. Thus, we decided to exclude them. In April 1981, we instituted a GENET user password checking system to further control community access, particularly in regard to industrial users.

4) Limited commitment of SUMEX staff resources -- The day to day management of the GENET community has been the responsibility of MOLGEN project personnel. SUMEX personnel have only contributed to developing system facilities to help manage GENET (guest and GENET password capabilities), assisted with technical communications problems, and advised in establishing GENET management policies consistent with AIM Executive Committee and SUMEX Principal Investigator resource policies. The total commitment of staff time has been on the order of 1-2 man-months.

GENET USER COMMUNITY

The GENET community consists of approximately 200 users from 63 research institutions. Of these 200 users, approximately 35 are consistently active users. That is, they log in, run programs, and interact with the MOLGEN members on an almost daily basis. Many of these users have made valuable contributions to our work. About 100 others are frequent, but not regular users. They log in only when they have a major analysis task to perform, which seems to be on the order of once a month.

The remaining users rarely use the system. They have logged in a few times, but for one reason or another they never become regular users of the system. Quite often this is because a lab group will settle on having one or two graduate students or post-doctoral associates become the "computer experts" of the group, and as a result, the computer use by the other people in the lab drops to an almost non-existent level. Unfortunately, an equally prevalent reason for users to stop using the GENET account is a lack of resource time. Probably the major complaint that we get from GENET users is concerning the lack of compute time and availability of the system. One account just is not enough for 200 people to share, especially when it is

ACCOUNT: GENET (February 1980 - March 1981)

MONTH/ YEAR	CPU HOURS	TOTAL CONNECT HOURS	TYMNET CONNECT HOURS	GENET % OF SUMEX TYMNET USE	FILE PAGES
Feb/80	3.23	32.72	18.88	2.0%	57
Mar/80	1.28	51.57	12.80	1.4	95
Apr/80	8.37	117.87	51.73	5.4	209
May/80	9.20	104.46	66.65	8.0	166
Jun/80	11.08	188.35	118.03	11.7	253
Jul/80	19.21	342.87	189.00	18.2	231
Aug/80	18.71	257.23	188.53	18.2	367
Sep/80	57.32	409.83	254.53	28.5	626
Oct/80	36.47	348.66	211.95	23.3	920
Nov/80	82.90	648.56	308.40	31.1	1133
Dec/80	19.86	295.85	188.67	22.8	1110
Jan/81	48.00	747.91	277.30	27.2	996
Feb/81	22.58	265.39	163.55	16.1	962
Mar/81	29.73	613.74	313.57	25.0	982

GENET USERS BY RESEARCH INSTITUTION

The following is list of GENET users by affiliation. The accompanying texts are excerpts from messages sent to MOLGEN project members concerning how the particular user plans to use or has used the GENET facility. For some universities we have complete lists of all professors, post-doctoral candidates, and graduate students using GENET. However, for many we have just a list of the most active users from the institution.

ALBERT EINSTEIN COLLEGE OF MEDICINE

Perry Nisen - Mol. Biol.
E. Benz - DBC
D. Brown - DBC
B. Birshtein - Cell Biol.
S. Tilley - Cell Biol.
G. Childs - Genetics
John Sninsky (EXO-MOLGEN user)
Bob Wydro
(3 additional graduate students)

I have been using the sequence analysis programs to identify sequence homologies between prokaryotic transposable DNA elements (i.e. Insertion Sequences and Transposable antibiotic resistance elements) and with E.coli RNA sequences. I have also been comparing some eukaryotic transposable sequences with the prokaryotic elements.